

Perspective Functions: Properties, Constructions, and Examples*

Patrick L. Combettes

North Carolina State University
Department of Mathematics
Raleigh, NC 27695-8205, USA
plc@math.ncsu.edu

Abstract

Many functions encountered in applied mathematics and in statistical data analysis can be expressed in terms of perspective functions. One of the earliest examples is the Fisher information, which appeared in statistics in the 1920s. We analyze various algebraic and convex-analytical properties of perspective functions and provide general schemes to construct lower semicontinuous convex functions from them. Several new examples are presented and existing instances are featured as special cases.

*Contact author: P. L. Combettes, plc@math.ncsu.edu, phone: +1 (919) 515 2671.

1 Introduction

Let \mathcal{G} be a real Hilbert space and let $\varphi: \mathcal{G} \rightarrow]-\infty, +\infty]$ be a convex function. The perspective function of φ is

$$\mathcal{P}_\varphi: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \eta\varphi(y/\eta), & \text{if } \eta > 0; \\ +\infty, & \text{otherwise.} \end{cases} \quad (1.1)$$

The properties of \mathcal{P}_φ were first investigated in [45], where it was shown in particular that \mathcal{P}_φ is convex (see also [6, 24, 29]). Special cases of the construction (1.1) arise in various areas of applied mathematics and data analysis. One of the oldest instances involving perspective functions is the Fisher information of a differentiable probability density $x: \mathbb{R}^N \rightarrow]0, +\infty[$, that is,

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt, \quad (1.2)$$

where $\|\cdot\|_2$ is the standard Euclidean norm on \mathbb{R}^N . This notion, which dates back to the work of Fisher in statistics [25], has found applications in many contexts, e.g., [9, 13, 14, 27, 41, 47]. More generally, (1.1) can be used to construct convex integrands of integral functionals such as

$$\int_{\mathbb{R}^N} \mathcal{P}_\varphi(x(t), y(t)) dt = \int_{\mathbb{R}^N} x(t) \varphi\left(\frac{y(t)}{x(t)}\right) dt, \quad (1.3)$$

where $x: \mathbb{R}^N \rightarrow]0, +\infty[$ and $y: \mathbb{R}^N \rightarrow \mathcal{G}$. In the case when $N = 1$ and $\mathcal{G} = \mathbb{R}$, it corresponds to a notion of φ -divergence which can be traced back to [2, 23] and that has been used extensively in information theory, statistics, signal processing, and pattern recognition [4, 10, 35, 44]; see also [7, 30] for a discussion of discrete counterparts. In the case when $\mathcal{G} = \mathbb{R}^{N \times N}$, $y = \nabla x$, and $\varphi = \|\cdot\|_2^2$, one recovers (1.2). Furthermore, choosing $\varphi = \|\cdot\|_2^p$ with $p \in]1, +\infty[$ provides the extension of the Fisher information (1.2) found in [12] in the case when $N = 1$. Instances of perspective functions can also be identified in robust estimation [32, Section 7.7] (see also [39, 42] for recent developments), transportation theory [8, 18, 26, 43], sparse regression [11, 34], control theory [33, 37], mixed-integer programming [28], computer vision [48], disjunctive programming [19], game theory [1], and machine learning [36].

Although perspective functions appear explicitly or implicitly in an increasing number of diverse research areas, little effort has been dedicated to the systematic study of their properties, especially in general Hilbert spaces. It is the goal of the present paper to propose such an investigation, with a special focus on the construction of lower semicontinuous convex functions around perspective functions. As is well known, these two properties are of paramount importance in the modeling, analysis, and numerical solution of variational problems. Section 2 focuses on algebraic and convex-analytical properties. In Section 3, several constructions of lower semicontinuous convex perspective functions are provided. Section 4 discusses the construction of lower semicontinuous convex functions based on perspective functions. Finally, integral functions with perspective function-based integrands are studied in Section 5. Many of the functions we propose are new and suggest new problem formulations in various applications areas. In particular, our results are exploited in the companion paper [22], which investigates the proximity operator of perspective functions and explores new models and algorithms in high-dimensional statistics.

Notation. Throughout, \mathcal{H} and \mathcal{G} are real Hilbert spaces and $\mathcal{H} \oplus \mathcal{G}$ denotes their Hilbert direct sum. The closed ball with center $x \in \mathcal{H}$ and radius $\rho \in]0, +\infty[$ in \mathcal{H} is denoted by $B(x; \rho)$. $\Gamma_0(\mathcal{H})$ is the class of lower semicontinuous convex functions $f: \mathcal{H} \rightarrow]-\infty, +\infty]$ such that $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \emptyset$. Let $f \in \Gamma_0(\mathcal{H})$. Then f^* denotes the conjugate of f , $\text{epi } f$ the epigraph of f , $\text{rec } f$ the recession function of f , and ∂f the subdifferential of f . Let C be a subset of \mathcal{H} . Then ι_C is the indicator function of C , d_C the distance function to C , $\text{rec } C$ the recession cone of C , and σ_C the support function of C . For background on convex analysis, see [6, 45, 49].

2 Properties of perspective functions

In this section we study various properties of perspective functions. We start our discussion by noting that, if $\varphi \in \Gamma_0(\mathcal{G})$, the construction (1.1) does not necessarily produce a lower semicontinuous function. For this reason, we shall use the following variant, first proposed in [45] for $\mathcal{G} = \mathbb{R}^N$.

Definition 2.1 Let $\varphi \in \Gamma_0(\mathcal{G})$ and let $\text{rec } \varphi$ be its recession function, i.e., given any $z \in \text{dom } \varphi$,

$$(\forall y \in \mathcal{G}) \quad (\text{rec } \varphi)(y) = \sup_{x \in \text{dom } \varphi} (\varphi(x + y) - \varphi(x)) = \lim_{\alpha \rightarrow +\infty} \frac{\varphi(z + \alpha y) - \varphi(z)}{\alpha} = \sigma_{\text{dom } \varphi^*}(y). \quad (2.1)$$

The lower semicontinuous envelope of the perspective of φ is

$$\tilde{\varphi}: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \eta \varphi(y/\eta), & \text{if } \eta > 0; \\ (\text{rec } \varphi)(y), & \text{if } \eta = 0; \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.2)$$

For simplicity, $\tilde{\varphi}$ is called the *perspective* of φ .

The following result captures basic topological and convex analytical properties of the perspective function (2.2).

Proposition 2.2 Let $\varphi \in \Gamma_0(\mathcal{G})$. Then the following hold:

- (i) $\tilde{\varphi}$ is positively homogeneous.
- (ii) $\tilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.
- (iii) $\tilde{\varphi}$ is sublinear.
- (iv) Let $C = \{(\mu, u) \in \mathbb{R} \times \mathcal{G} \mid \mu + \varphi^*(u) \leq 0\}$. Then $(\tilde{\varphi})^* = \iota_C$ and $\tilde{\varphi} = \sigma_C$.
- (v) Let $\eta \in \mathbb{R}$ and $y \in \mathcal{G}$. Then

$$\partial \tilde{\varphi}(\eta, y) = \begin{cases} \{(\varphi(y/\eta) - \langle y \mid u \rangle / \eta, u) \mid u \in \partial \varphi(y/\eta)\}, & \text{if } \eta > 0; \\ \{(\mu, u) \in C \mid \sigma_{\text{dom } \varphi^*}(y) = \langle u \mid y \rangle\}, & \text{if } \eta = 0 \text{ and } y \neq 0; \\ C, & \text{if } \eta = 0 \text{ and } y = 0; \\ \emptyset, & \text{if } \eta < 0. \end{cases} \quad (2.3)$$

Proof. (i): This follows from (2.1) and (2.2).

(ii): Set $D = \{1\} \times \text{epi } \varphi$ and $g = \mathcal{P}_\varphi$, and let $z \in \text{dom } \varphi$. Then $(1, z) \in \text{dom } g$. On the other hand, $\text{epi } g = \text{cone } D$ is convex and g is therefore a proper convex function. Let us denote by \check{g} the largest lower semicontinuous convex function majorized by g . To show that $\tilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$, it is enough to show that

$$\tilde{\varphi} = \check{g}. \quad (2.4)$$

This can be done using the following argument due to H. H. Bauschke. Since $(0, 0) \notin D$ and $\text{repepi } \varphi = \text{epi rec } \varphi$ [49, p. 74], it follows from [6, Theorem 9.9, Corollary 6.52, and Lemma 1.6(ii)] that $\text{epi } \check{g} = \overline{\text{epi } f} = \overline{\text{cone } D} = (\text{cone } D) \cup (\text{rec } D) = (\text{epi } f) \cup (\{0\} \times \text{rec epi } \varphi) = (\text{epi } f) \cup (\{0\} \times \text{epi rec } \varphi) = (\text{epi } f) \cup \text{epi } (\iota_{\{0\}} \oplus \text{rec } \varphi) = \text{epi min}\{f, \iota_{\{0\}} \oplus \text{rec } \varphi\} = \text{epi } \tilde{\varphi}$.

(iii): This follows from (i) and (ii).

(iv): Set $g = \mathcal{P}_\varphi$. Then $g^* = \iota_C$ [6, Example 13.8]. Hence, we derive from (2.4) and [6, Proposition 13.14] that $(\tilde{\varphi})^* = (\check{g})^* = g^* = \iota_C$. In turn, (ii) and [6, Corollary 13.33] yield $\tilde{\varphi} = (\tilde{\varphi})^{**} = \iota_C^* = \sigma_C$.

(v): It follows from the Fenchel-Young identity ([6, Proposition 16.13]) and (iv) that

$$\begin{aligned} (\mu, u) \in \partial \tilde{\varphi}(\eta, y) &\Leftrightarrow \tilde{\varphi}(\eta, y) + (\tilde{\varphi})^*(\mu, u) = \eta\mu + \langle y \mid u \rangle \\ &\Leftrightarrow \tilde{\varphi}(\eta, y) = \eta\mu + \langle y \mid u \rangle \text{ and } \mu + \varphi^*(u) \leq 0. \end{aligned} \quad (2.5)$$

We consider three cases.

- $\eta < 0$: Then (2.2) and (2.5) yield $\partial \tilde{\varphi}(\eta, y) = \emptyset$.
- $\eta = 0$: We deduce from (2.5), (2.2), and (2.1) that

$$\begin{aligned} (\mu, u) \in \partial \tilde{\varphi}(\eta, y) &\Leftrightarrow (\text{rec } \varphi)(y) = \langle y \mid u \rangle \text{ and } \mu + \varphi^*(u) \leq 0 \\ &\Leftrightarrow \sigma_{\text{dom } \varphi^*}(y) = \langle y \mid u \rangle \text{ and } (\mu, u) \in C. \end{aligned} \quad (2.6)$$

Since $\sigma_{\text{dom } \varphi^*}(0) = 0 = \langle 0 \mid u \rangle$, we obtain the desired results.

- $\eta > 0$: Using successively (2.5), (2.2), the Fenchel-Young inequality ([6, Proposition 13.13]), and the Fenchel-Young identity, we obtain

$$\begin{aligned} (\mu, u) \in \partial \tilde{\varphi}(\eta, y) &\Leftrightarrow \mu = \varphi(y/\eta) - \langle y \mid u \rangle / \eta \text{ and } \varphi(y/\eta) + \varphi^*(u) \leq \langle y/\eta \mid u \rangle \\ &\Leftrightarrow \mu = \varphi(y/\eta) - \langle y \mid u \rangle / \eta \text{ and } \varphi(y/\eta) + \varphi^*(u) = \langle y/\eta \mid u \rangle \\ &\Leftrightarrow \mu = \varphi(y/\eta) - \langle y \mid u \rangle / \eta \text{ and } u \in \partial \varphi(y/\eta). \end{aligned} \quad (2.7)$$

We have thus proved (2.3). \square

Remark 2.3 Some of the results of Proposition 2.2 have already been obtained in the when $\mathcal{G} = \mathbb{R}^N$ with different tools. Thus, items (ii) and (iv) can be found in [45], and the case $\eta > 0$ of (v) appears in [19, Proposition 4].

As will be seen in [22], (2.3) is instrumental in computing the proximity operator of a perspective function. Here is an important refinement.

Corollary 2.4 *Let $\varphi \in \Gamma_0(\mathcal{G})$ and denote by $\text{bar dom } \varphi^*$ the barrier cone of $\text{dom } \varphi^*$. Let $\eta \in \mathbb{R}$, let $y \in \mathcal{G}$, and suppose that one of the following holds:*

- (i) $y \notin \text{bar dom } \varphi^*$.
- (ii) $\text{dom } \varphi^*$ is open.
- (iii) $\text{dom } \varphi^* = \mathcal{G}$.
- (iv) φ is supercoercive: $\lim_{\|y\| \rightarrow +\infty} \varphi(y)/\|y\| = +\infty$.
- (v) For every $u \in \mathcal{G}$, $\varphi - \langle \cdot | u \rangle$ is coercive.

Then

$$\partial \tilde{\varphi}(\eta, y) = \begin{cases} \{(\varphi(y/\eta) - \langle y | u \rangle / \eta, u) \mid u \in \partial \varphi(y/\eta)\}, & \text{if } \eta > 0; \\ C, & \text{if } \eta = 0 \text{ and } y = 0; \\ \emptyset, & \text{otherwise.} \end{cases} \quad (2.8)$$

Proof. In view of Proposition 2.2(v), it suffices to suppose that $y \neq 0$ and to show that

$$D = \{(\mu, u) \in \mathbb{R} \times \mathcal{G} \mid \mu + \varphi^*(u) \leq 0 \text{ and } \sigma_{\text{dom } \varphi^*}(y) = \langle u | y \rangle\} = \emptyset. \quad (2.9)$$

Now denote by $\text{spts dom } \varphi^*$ the set of support points of $\text{dom } \varphi^*$. Then

$$D = \{(\mu, u) \in \mathbb{R} \times (\text{spts dom } \varphi^*) \mid \mu + \varphi^*(u) \leq 0 \text{ and } \sigma_{\text{dom } \varphi^*}(y) = \langle u | y \rangle\}. \quad (2.10)$$

(i): We have $\sigma_{\text{dom } \varphi^*}(y) = +\infty$ and therefore (2.9) yields $D = \emptyset$.

(ii): We have $\text{spts dom } \varphi^* = \emptyset$ and therefore (2.10) yields $D = \emptyset$.

(iii) \Rightarrow (ii): Clear.

(iv) \Rightarrow (iii): [6, Proposition 14.15].

(v) \Rightarrow (iii): Let $u \in \mathcal{G}$. Then by the Moreau-Rockafellar theorem ([6, Theorem 14.17]), $\varphi - \langle \cdot | u \rangle$ is coercive if and only if $u \in \text{int dom } \varphi^*$. Hence $\mathcal{G} \subset \text{int dom } \varphi^*$. \square

Next, we provide an example of a perspective function $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ such that $g|_{\text{dom } g}$ is discontinuous.

Example 2.5 Suppose that $\mathcal{G} \neq \{0\}$, let $p \in]1, +\infty[$, and set

$$g: \mathbb{R} \oplus \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \|y\|^p / \eta^{p-1}, & \text{if } \eta > 0; \\ 0, & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.11)$$

Then $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ and $g|_{\text{dom } g}$ is not continuous at $(0, 0)$. Indeed, set $\varphi = \|\cdot\|^p$. Then φ is a supercoercive function in $\Gamma_0(\mathcal{G})$, and it thus follows from (2.1) that $\text{rec } \varphi = \iota_{\{0\}}$. Hence (2.11) coincides with (2.2) and the first claim is therefore an application of Proposition 2.2(ii) with $g = \tilde{\varphi}$. Now set $y = (0, 0) \in \mathbb{R} \times \mathcal{G}$, let $v \in \mathcal{G}$ be such that $\|v\| = 1$, fix a sequence $(\alpha_n)_{n \in \mathbb{N}}$ in $]0, +\infty[$ such that $\alpha_n \downarrow 0$, and set $(\forall n \in \mathbb{N}) \ y_n = (\alpha_n^{p/(p-1)}, \alpha_n v)$. Then $(y_n)_{n \in \mathbb{N}}$ lies in $\text{dom } g$ and $y_n \rightarrow y$, but $\lim g(y_n) = 1 \neq 0 = g(y)$.

We now turn to some algebraic properties.

Proposition 2.6 *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then the following hold:*

- (i) *Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\text{dom } \varphi \cap \text{dom } \psi \neq \emptyset$, and let $\lambda \in]0, +\infty[$. Then $[\lambda\varphi + \psi]^\sim = \lambda\tilde{\varphi} + \tilde{\psi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*
- (ii) *Let $\Lambda: \mathcal{H} \rightarrow \mathcal{G}$ be linear, bounded, and such that $\text{ran } \Lambda \cap \text{dom } \varphi \neq \emptyset$. Set $\tilde{\Lambda}: \mathbb{R} \oplus \mathcal{H} \rightarrow \mathbb{R} \oplus \mathcal{G}: (\xi, x) \mapsto (\xi, \Lambda x)$. Then $[\varphi \circ \Lambda]^\sim = \tilde{\varphi} \circ \tilde{\Lambda} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*

Proof. (i): We have $\text{dom } (\varphi + \psi) \neq \emptyset$. Hence $\varphi + \psi \in \Gamma_0(\mathcal{G})$ and (2.1) implies that $\text{rec } (\lambda\varphi + \psi) = \lambda \text{rec } \varphi + \text{rec } \psi$. The claim therefore follows from (2.2) and Proposition 2.2(ii).

(ii): Let $\xi \in \mathbb{R}$ and $x \in \mathcal{H}$. If $\xi > 0$, then $[\varphi \circ \Lambda]^\sim(\xi, x) = \xi(\varphi \circ \Lambda)(x/\xi) = \xi\varphi(\Lambda x/\xi) = (\tilde{\varphi} \circ \tilde{\Lambda})(\xi, x)$. Furthermore, we have $\text{dom } (\varphi \circ \Lambda) \neq \emptyset$. Hence, $\varphi \circ \Lambda \in \Gamma_0(\mathcal{H})$ and (2.1) yields $\text{rec } (\varphi \circ \Lambda) = (\text{rec } \varphi) \circ \Lambda$. Hence, we derive from (2.2) that $[\varphi \circ \Lambda]^\sim(0, x) = \text{rec } (\varphi \circ \Lambda)(x) = (\text{rec } \varphi)(\Lambda x) = (\tilde{\varphi} \circ \tilde{\Lambda})(0, x)$. Finally, if $\xi < 0$, then $[\varphi \circ \Lambda]^\sim(\xi, x) = +\infty = (\tilde{\varphi} \circ \tilde{\Lambda})(\xi, x)$. Altogether, the conclusion follows from Proposition 2.2(ii). \square

Corollary 2.7 *Let $\psi \in \Gamma_0(\mathcal{G})$ and let C be a closed convex subset of \mathcal{G} such that $C \cap \text{dom } \psi \neq \emptyset$. Set*

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty]: (\eta, y) \mapsto \begin{cases} \eta\psi(y/\eta), & \text{if } \eta > 0 \text{ and } y \in \eta(C \cap \text{dom } \psi); \\ (\text{rec } \psi)(y), & \text{if } \eta = 0 \text{ and } y \in \text{rec } C; \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.12)$$

Then $g \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Proof. This is an application of Proposition 2.6(i) with $\lambda = 1$ and $\varphi = \iota_C$. Indeed, in this setting, $\text{rec } (\varphi + \psi) = \text{rec } \iota_C + \text{rec } \psi = \iota_{\text{rec } C} + \text{rec } \psi$ and (2.12) yields $g = [\iota_C + \psi]^\sim$. \square

Corollary 2.8 *Let $\varphi \in \Gamma_0(\mathcal{G})$, let $\psi \in \Gamma_0(\mathcal{G})$ be a positively homogeneous function such that $\text{dom } \varphi \cap \text{dom } \psi \neq \emptyset$, and let $\delta \in \mathbb{R}$. Then $[\varphi + \psi + \delta]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$ and*

$$(\forall \eta \in \mathbb{R})(\forall y \in \mathcal{G}) \quad [\varphi + \psi + \delta]^\sim(\eta, y) = \tilde{\varphi}(\eta, y) + \psi(y) + \delta\eta. \quad (2.13)$$

Proof. This follows from (2.2) and Proposition 2.6(i) since $\text{rec } (\varphi + \psi + \delta) = (\text{rec } \varphi) + (\text{rec } \psi) = (\text{rec } \varphi) + \psi$. \square

Corollary 2.9 *Let $\varphi \in \Gamma_0(\mathcal{G})$. Then $(\forall (\zeta, \eta) \in \mathbb{R}^2)(\forall y \in \mathcal{G}) \ \tilde{\tilde{\varphi}}(\zeta, \eta, y) = \tilde{\varphi}(\eta, y)$.*

Proof. By Proposition 2.2(i)–(ii), $\tilde{\varphi}$ is a positively homogeneous function in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Hence the claim follows from Corollary 2.8. \square

Proposition 2.10 *Let I be a finite set and let $\eta \in \mathbb{R}$. For every $i \in I$, let \mathcal{G}_i be a real Hilbert space, let $\varphi_i \in \Gamma_0(\mathcal{G}_i)$, and let $y_i \in \mathcal{G}_i$. Set $\bigoplus_{i \in I} \varphi_i: (z_i)_{i \in I} \mapsto \sum_{i \in I} \varphi_i(z_i)$. Then*

$$\left(\bigoplus_{i \in I} \varphi_i \right)^\sim (\eta, (y_i)_{i \in I}) = \left(\bigoplus_{i \in I} \tilde{\varphi}_i \right) ((\eta, y_i))_{i \in I}. \quad (2.14)$$

Proof. Suppose that $\eta > 0$. Then

$$\left(\bigoplus_{i \in I} \varphi_i \right)^\sim (\eta, (y_i)_{i \in I}) = \eta \left(\bigoplus_{i \in I} \varphi_i \right) (y_i/\eta)_{i \in I} = \sum_{i \in I} \eta \varphi_i(y_i/\eta) = \left(\bigoplus_{i \in I} \tilde{\varphi}_i \right) ((\eta, y_i))_{i \in I}. \quad (2.15)$$

Now suppose that $\eta = 0$. Then (2.1) implies that $\text{rec } \bigoplus_{i \in I} \varphi_i = \bigoplus_{i \in I} \text{rec } \varphi_i$ and (2.14) follows. Finally, if $\eta < 0$, then both sides of (2.14) are equal to $+\infty$. \square

Perspective functions can be used to provide examples of nonintuitive behaviors for minimizing sequences in optimization problems.

Example 2.11 Suppose that $\mathcal{G} = \mathbb{R}$. Then Proposition 2.2(ii) asserts that the function

$$g = [|\cdot|^2]^\sim: \mathbb{R}^2 \rightarrow]-\infty, +\infty]: (\xi_1, \xi_2) \mapsto \begin{cases} \xi_2^2/\xi_1, & \text{if } \xi_1 > 0; \\ 0, & \text{if } \xi_1 = \xi_2 = 0; \\ +\infty, & \text{otherwise} \end{cases} \quad (2.16)$$

belongs to $\Gamma_0(\mathbb{R}^2)$. Moreover, $\text{Argmin } g = [0, +\infty[\times \{0\}$. Now let $p \in [1, +\infty[$ and set $(\forall n \in \mathbb{N})$ $x_n = ((n+1)^{p+2}, n+1)$. Then $(x_n)_{n \in \mathbb{N}}$ is a minimizing sequence of g since $g(x_n) - \min g(\mathbb{R}^2) = 1/(n+1)^p \downarrow 0$. However, $d_{\text{Argmin } g}(x_n) = n+1 \uparrow +\infty$. To sum up,

$$g(x_n) - \min g(\mathbb{R}^2) = O(1/n^p), \quad \text{while} \quad (\forall x \in \text{Argmin } g) \quad \|x_n - x\| \uparrow +\infty. \quad (2.17)$$

This illustrates the fact that, even if it induces a very good convergence rate of the objective values $(g(x_n))_{n \in \mathbb{N}}$, a minimizing sequence $(x_n)_{n \in \mathbb{N}}$ may have extremely poor properties in terms of actually approaching a solution to the underlying minimization problem.

3 Examples of perspective functions

Our first construction involves a difference of convex functions.

Corollary 3.1 *Let $\psi \in \Gamma_0(\mathcal{G})$ and let $\text{env } \psi: y \mapsto \inf_{x \in \mathcal{G}} (\psi(x) + \|y - x\|^2/2)$ be the Moreau envelope of ψ . Set*

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty]: (\eta, y) \mapsto \begin{cases} \frac{\|y\|^2}{2\eta} - \eta(\text{env } \psi)(y/\eta), & \text{if } \eta > 0; \\ \sigma_{\text{dom } \psi}(y), & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.1)$$

Then $g = [\text{env } (\psi^)]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.*

Proof. Set $q = \|\cdot\|^2/2$ and $\varphi = q - \text{env } \psi$, and let \square denote the infimal convolution operation. It follows from Moreau's decomposition [38] (see also [6, Theorem 14.3(i)]) that $\varphi = \text{env}(\psi^*) \in \Gamma_0(\mathcal{G})$. In addition, from basic convex analysis, $\varphi^* = (\psi^* \square q)^* = \psi^{**} + q = \psi + q$ and therefore (2.1) yields

$$\text{rec } \varphi = \sigma_{\text{dom } \varphi^*} = \sigma_{\text{dom } \psi}. \quad (3.2)$$

In view of (2.2) and Proposition 2.2(ii), we conclude that $g = \tilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. \square

Example 3.2 (generalized Huber function) Let C be a nonempty closed convex subset of \mathcal{G} and let P_C denote its projector. Upon setting $\psi = \iota_C$ in Corollary 3.1, we deduce that the function

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \langle y | P_C(y/\eta) \rangle - \frac{\eta \|P_C(y/\eta)\|^2}{2}, & \text{if } y \notin \eta C \text{ and } \eta > 0; \\ \frac{\|y\|^2}{2\eta}, & \text{if } y \in \eta C \text{ and } \eta > 0; \\ \sigma_C(y), & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0 \end{cases} \quad (3.3)$$

is in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. More precisely, $g = \tilde{\varphi}$, where $\varphi = \text{env}(\psi^*) = \text{env } \sigma_C$. Let us further specialize by taking $C = B(0; \rho)$ for some $\rho \in]0, +\infty[$. Then (3.3) reduces to

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \rho \|y\| - \frac{\eta \rho^2}{2}, & \text{if } \|y\| > \eta \rho \text{ and } \eta > 0; \\ \frac{\|y\|^2}{2\eta}, & \text{if } \|y\| \leq \eta \rho \text{ and } \eta > 0; \\ \rho \|y\|, & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.4)$$

To sum up, we infer from Corollary 3.1 that $g = \tilde{\varphi}$, where $\varphi = \text{env } \|\cdot\| = q - d_C^2/2$, that is

$$\varphi: \mathcal{G} \rightarrow]-\infty, +\infty] : y \mapsto \begin{cases} \rho \|y\| - \frac{\rho^2}{2}, & \text{if } \|y\| > \rho; \\ \frac{\|y\|^2}{2}, & \text{if } \|y\| \leq \rho. \end{cases} \quad (3.5)$$

In particular, if $\mathcal{G} = \mathbb{R}$, then φ is known as the Huber function. This function was introduced in [31] and it plays an important role in robust statistics and signal processing [32, 40], while its perspective function appears implicitly in robust regression problems [32, 42]. The fact that the Huber function on \mathbb{R} is the Moreau envelope of the absolute value function can already be found in [16]; see also [17]. On the other hand, if we specialize the perspective function (3.4) to the case when $\mathcal{G} = \mathbb{R}$ and $\rho = 1$, we obtain the function

$$g: \mathbb{R}^2 \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} |y| - \frac{\eta}{2}, & \text{if } |y| > \eta \text{ and } \eta > 0; \\ \frac{|y|^2}{2\eta}, & \text{if } |y| \leq \eta \text{ and } \eta > 0; \\ |y|, & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0, \end{cases} \quad (3.6)$$

which is used in computer vision [48], where it is called the bivariate Huber function.

We now consider a function that combines distance and support functions.

Example 3.3 (generalized Berhu function) Let C and D be nonempty closed convex subsets of \mathcal{G} , and let $\rho \in]0, +\infty[$. Then the function

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \frac{\eta d_C^2(y/\eta)}{2\rho} + \sigma_D(y), & \text{if } \eta > 0 \text{ and } y \notin \eta C; \\ \sigma_D(y), & \text{if } \eta > 0 \text{ and } y \in \eta C; \\ \sigma_D(y), & \text{if } \eta = 0 \text{ and } y \in \text{rec } C; \\ +\infty, & \text{otherwise} \end{cases} \quad (3.7)$$

is in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. To show this, set $q = \|\cdot\|^2/2$ and $\vartheta = \varphi + \psi$, where $\varphi = d_C^2/(2\rho)$ and $\psi = \sigma_D$. Then $\varphi \in \Gamma_0(\mathcal{G})$ and ψ is a positively homogeneous function in $\Gamma_0(\mathcal{G})$ such that $0 \in \text{dom } \varphi \cap \text{dom } \psi$. Furthermore, since $\varphi = \iota_C \square (q/\rho)$, we have $\varphi^* = \iota_C^* + (q/\rho)^* = \iota_C^* + \rho q$ and therefore $\text{dom } \varphi^* = \text{dom } \iota_C^*$. In turn $\text{rec } \varphi = \sigma_{\text{dom } \varphi^*} = \sigma_{\text{dom } \iota_C^*} = \text{rec } \iota_C = \iota_{\text{rec } C}$. Altogether $g = \widetilde{\vartheta}$, more specifically,

$$g = \left[\frac{d_C^2}{2\rho} + \sigma_D \right]^\sim \quad (3.8)$$

and the claim follows from Corollary 2.8. An especially interesting case is obtained when $C = B(0; \rho)$ and $D = B(0; 1)$. Then $\text{rec } C = \{0\}$, $\sigma_D = \|\cdot\|$, and (3.7) therefore becomes

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \frac{\|y\|^2 + \rho^2 \eta^2}{2\eta\rho}, & \text{if } \eta > 0 \text{ and } \|y\| > \eta\rho; \\ \|y\|, & \text{if } \eta > 0 \text{ and } \|y\| \leq \eta\rho; \\ 0, & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \quad (3.9)$$

As seen above, g is the perspective function of

$$\vartheta: \mathcal{G} \rightarrow]-\infty, +\infty] : y \mapsto \begin{cases} \frac{\|y\|^2 + \rho^2}{2\rho}, & \text{if } \|y\| > \rho; \\ \|y\|, & \text{if } \|y\| \leq \rho. \end{cases} \quad (3.10)$$

In the special case when $\mathcal{G} = \mathbb{R}$, ϑ arises in mechanics [3, 15] as well as in statistics [42], where it is called the Berhu (or reverse Huber) function. The reason for this terminology is that (3.5) exhibits a quadratic behavior on $B(0; \rho)$ and a sublinear behavior outside, while (3.10) exhibits a sublinear behavior on $B(0; \rho)$ and a quadratic behavior outside. Applications of the perspective of the Berhu function in robust regression can be found in [42].

We now turn to a type of function that is used in support vector machines and in computer vision.

Example 3.4 (generalized Vapnik loss function) Let $\psi: \mathcal{G} \rightarrow]-\infty, +\infty]$ be a proper lower semi-continuous positively homogeneous convex function, let $\varepsilon \in \mathbb{R}$, and set

$$g: \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \max\{\psi(y) - \varepsilon\eta, 0\}, & \text{if } \eta \geq 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.11)$$

Then we derive from (2.1), (2.2), and Proposition 2.2(ii) that

$$g = [\max\{\psi - \varepsilon, 0\}]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G}). \quad (3.12)$$

Now consider the special case when $\psi = \|\cdot\|$ and $\varepsilon \in]0, +\infty[$. Then

$$g = [\max\{\|\cdot\| - \varepsilon, 0\}]^\sim : \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} d_{B(0; \varepsilon \eta)}(y), & \text{if } \eta > 0; \\ \|y\|, & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.13)$$

A special case of this function appears in the context of computer vision in [48]. When $\mathcal{G} = \mathbb{R}$, the function $\max\{\|\cdot\| - \varepsilon, 0\}$ is known as Vapnik's ε -insensitive loss function and it is employed in the area of support vector machine [46].

Our next construction involves a mix of positively homogeneous and norm-like functions.

Example 3.5 Let $\psi : \mathcal{G} \rightarrow [0, +\infty]$ be a proper lower semicontinuous positively homogeneous convex function, let $\delta \in \mathbb{R}$, let $\rho \in [0, +\infty[$, let $p \in [1, +\infty[$, and set

$$g : \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \delta\eta + |\rho\eta^p + \psi^p(y)|^{1/p}, & \text{if } \eta \geq 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.14)$$

Then $g = [\delta + |\rho + \psi^p|^{1/p}]^\sim \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. Indeed, set $\varphi = \delta + |\rho + \psi^p|^{1/p}$. We first observe that the function $\phi : \mathbb{R} \rightarrow \mathbb{R} : t \mapsto |\rho + |t|^p|^{1/p}$ is increasing on $[0, +\infty[$, continuous, and convex. Hence, upon extending ϕ by setting $\phi(+\infty) = +\infty$, we see that $\varphi = \delta + \phi \circ \psi \in \Gamma_0(\mathcal{G})$. Next, we derive from (2.1) and the positive homogeneity of ψ that $\text{rec } \varphi = \text{rec } |\rho + \psi^p|^{1/p} = \text{rec } \psi = \psi$. Thus, if $\eta = 0$, $(\forall y \in \mathcal{G}) \ |\rho\eta^p + \psi^p(y)|^{1/p} + \delta\eta = \psi(y) = (\text{rec } \varphi)(y)$. Altogether, in view of Proposition 2.2(ii) and (2.2), the assertion is proved. Let us now consider some special cases of this perspective function.

(i) Set $\psi = \|\cdot\|$, $\delta = 0$, and $p = 2$. Then (3.14) leads to the perspective function

$$g : \mathbb{R} \times \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \sqrt{\rho\eta^2 + \|y\|^2}, & \text{if } \eta \geq 0; \\ +\infty, & \text{if } \eta < 0. \end{cases} \quad (3.15)$$

(ii) Let D be a nonempty closed convex cone in \mathcal{G} , let $\delta = 0$, let $\rho = 1$, and let $\|\cdot\|$ be a norm on \mathcal{G} . Set $\mathcal{H} = \mathbb{R} \oplus \mathcal{G}$, $K = [0, +\infty[\times D$, and $\psi = \|\cdot\| + \iota_D$. Define a norm on \mathcal{H} by $\|\cdot\|_p : (\eta, y) \mapsto (|\eta|^p + \|\eta y\|^p)^{1/p}$. Then (3.14) yields $g = \|\cdot\|_p + \iota_K$, i.e.,

$$g : \mathcal{H} \rightarrow]-\infty, +\infty] : z \mapsto \begin{cases} \|z\|_p, & \text{if } z \in K; \\ +\infty, & \text{if } z \notin K. \end{cases} \quad (3.16)$$

(iii) Consider the following setting in (ii): $N \geq 2$ is an integer, $\mathcal{G} = \mathbb{R}^{N-1}$, $\|\cdot\|$ is the ℓ^p norm on \mathbb{R}^{N-1} , $D = [0, +\infty[^{N-1}$, and $K = [0, +\infty[^N$. Then, if $\|\cdot\|_p$ denotes the ℓ^p norm on \mathbb{R}^N , the corresponding perspective function (3.14) is

$$g : \mathbb{R}^N \rightarrow]-\infty, +\infty] : z \mapsto \begin{cases} \|z\|_p, & \text{if } z \in [0, +\infty[^N; \\ +\infty, & \text{if } z \notin [0, +\infty[^N. \end{cases} \quad (3.17)$$

The last example of this section extends constructions found in robust estimation and in machine learning.

Example 3.6 Let $\phi \in \Gamma_0(\mathbb{R})$ be an even function, let $v \in \mathcal{G}$, and let $\delta \in \mathbb{R}$. Then the function

$$g: \mathbb{R} \oplus \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \eta\phi(\|y\|/\eta) + \langle y | v \rangle + \delta\eta, & \text{if } \eta > 0; \\ (\text{rec } \phi)(\|y\|) + \langle y | v \rangle, & \text{if } \eta = 0; \\ +\infty, & \text{if } \eta < 0 \end{cases} \quad (3.18)$$

is in $\Gamma_0(\mathbb{R} \oplus \mathcal{G})$. More precisely, $g = [\varphi + \psi]^\sim$, where $\varphi = \phi \circ \|\cdot\| + \delta \in \Gamma_0(\mathcal{G})$ and $\phi = \langle \cdot | v \rangle \in \Gamma_0(\mathcal{G})$ is positively homogeneous. Note also that, since $\phi(0) = 0$, we have $0 \in \text{dom } \varphi \cap \text{dom } \psi$. Thus, the claim follows from Corollary 2.8. Now assume further that $\text{dom } \phi^* = \mathbb{R}$. Then we derive from [6, Example 13.7 and Proposition 13.20(iii)] that $\varphi^* = \phi^* \circ \|\cdot - v\| - \delta$. On the other hand, since $\text{dom } \phi^* = \mathbb{R}$, [5, Theorem 3.4] implies that $\phi^{**} = \phi$ is supercoercive and, therefore, that φ is likewise. In turn, we derive from (2.1) that $\text{rec } \varphi = \iota_{\{0\}}$, which allows us to rewrite (3.18) as

$$g: \mathbb{R} \oplus \mathcal{G} \rightarrow]-\infty, +\infty] : (\eta, y) \mapsto \begin{cases} \eta\phi(\|y\|/\eta) + \langle y | v \rangle + \delta\eta, & \text{if } \eta > 0; \\ 0 & \text{if } \eta = 0 \text{ and } y = 0; \\ +\infty, & \text{otherwise.} \end{cases} \quad (3.19)$$

In particular, when $\mathcal{G} = \mathbb{R}$, $\phi = |\cdot|^2$, and $v = 0$, (3.19) has been used in robust estimation [32] and in machine learning [36].

4 Composite perspective functions

We now describe various constructions of lower semicontinuous convex functions based on perspective functions. Our first result is based on the composition of the perspective of a convex function with an affine operator.

Proposition 4.1 Let $L: \mathcal{H} \rightarrow \mathcal{G}$ be linear and bounded, let $\varphi \in \Gamma_0(\mathcal{G})$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and set

$$f: \mathcal{H} \rightarrow]-\infty, +\infty] : x \mapsto \begin{cases} (\langle x | u \rangle - \rho)\varphi\left(\frac{Lx - r}{\langle x | u \rangle - \rho}\right), & \text{if } \langle x | u \rangle > \rho; \\ (\text{rec } \varphi)(Lx - r), & \text{if } \langle x | u \rangle = \rho; \\ +\infty, & \text{if } \langle x | u \rangle < \rho. \end{cases} \quad (4.1)$$

Suppose that there exists $z \in \mathcal{H}$ such that $Lz \in r + (\langle z | u \rangle - \rho)\text{dom } \varphi$ and $\langle z | u \rangle \geq 0$, and set $A: \mathcal{H} \rightarrow \mathbb{R} \oplus \mathcal{G}: x \mapsto (\langle x | u \rangle - \rho, Lx - r)$. Then $f = \tilde{\varphi} \circ A \in \Gamma_0(\mathcal{H})$.

Proof. By construction, A is a continuous affine operator, while $\tilde{\varphi} \in \Gamma_0(\mathcal{G})$ by Proposition 2.2(ii). Therefore $f = \tilde{\varphi} \circ A$ is lower semicontinuous and convex. Finally, the existence of z guarantees that f is proper. \square

Proposition 4.1 is an effective device for constructing lower semicontinuous convex functions from a perspective function. In particular, various examples in $\Gamma_0(\mathcal{H})$ can be systematically created

by composing a specific perspective function $\tilde{\varphi}$ from Section 3 for some $\varphi \in \Gamma_0(\mathcal{G})$ with a continuous affine operator $A: \mathcal{H} \rightarrow \mathbb{R} \oplus \mathcal{G}$ and, possibly, a suitable convexity preserving operation. Here is an illustration (see also Proposition 5.2).

Example 4.2 Let $L: \mathcal{H} \rightarrow \mathcal{G}$ be linear and bounded, let $||| \cdot |||$ be a norm on \mathcal{G} such that, for some $\chi \in]0, +\infty[$, $||| \cdot ||| \geq \chi \| \cdot \|$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and let q and s be in $]1, +\infty[$. Set

$$h: \mathcal{H} \rightarrow]-\infty, +\infty]: x \mapsto \begin{cases} \frac{|||Lx - r|||^{qs}}{|\langle x | u \rangle - \rho|^{(q-1)s}}, & \text{if } \langle x | u \rangle > \rho; \\ 0, & \text{if } Lx = r \text{ and } \langle x | u \rangle = \rho; \\ +\infty, & \text{otherwise.} \end{cases} \quad (4.2)$$

Then $h \in \Gamma_0(\mathcal{H})$.

Proof. Set $\varphi = ||| \cdot |||^q$. Then $\text{dom } \varphi = \mathcal{G}$. In addition, $\varphi(y)/\|y\| \geq \chi |||y|||^q / \|y\| \rightarrow +\infty$ as $\|y\| \rightarrow +\infty$ and therefore (2.1) implies that $\text{rec } \varphi = \iota_{\{0\}}$. Thus, (4.1) becomes

$$f: \mathcal{H} \rightarrow]-\infty, +\infty]: x \mapsto \begin{cases} \frac{|||Lx - r|||^q}{|\langle x | u \rangle - \rho|^{q-1}}, & \text{if } \langle x | u \rangle > \rho; \\ 0, & \text{if } Lx = r \text{ and } \langle x | u \rangle = \rho; \\ +\infty, & \text{otherwise,} \end{cases} \quad (4.3)$$

and Proposition 4.1 asserts that $f \in \Gamma_0(\mathcal{H})$. Now let $\phi = | \cdot |^s$. Then ϕ is increasing on $[0, +\infty] = \text{ran } f$, continuous, and convex. Hence it follows from [20, Proposition II.8.4] and [6, Proposition 8.19] that $h = \phi \circ f \in \Gamma_0(\mathcal{H})$. \square

Example 4.3 Let (Ω, \mathcal{F}, P) be a probability space, let $\mathcal{H} = L^2(\Omega, \mathcal{F}, P)$ be the associated Hilbert space of square-integrable random variables, let $p \in]1, 2]$, and let q and s be in $]1, +\infty[$. Set

$$h: \mathcal{H} \rightarrow]-\infty, +\infty]: X \mapsto \begin{cases} \frac{E^{qs/p}|X|^p}{E^{(q-1)s}X}, & \text{if } EX > 0; \\ 0, & \text{if } X = 0 \text{ a.s.}; \\ +\infty, & \text{otherwise.} \end{cases} \quad (4.4)$$

Then $h \in \Gamma_0(\mathcal{H})$. To see this, it suffices to apply Example 4.2 with $\mu = P$, $||| \cdot |||: X \mapsto E^{1/p}|X|^p$, $u = 1$ a.s., $r = 0$, and $\rho = 0$. The convexity property of (4.4) leads to potentially interesting moment inequalities in probability.

A special case of the function arising in the following result appears in [1] in the context of game theory. It involves the marginal of a perspective function.

Proposition 4.4 Let $\varphi \in \Gamma_0(\mathcal{G})$ and let K be a nonempty closed bounded interval in $[0, +\infty[$. Define

$$g: \mathcal{G} \rightarrow \mathbb{R}: y \mapsto \inf_{\eta \in K} \tilde{\varphi}(\eta, y). \quad (4.5)$$

Then $g \in \Gamma_0(\mathcal{G})$.

Proof. Proposition 2.2(ii) asserts that $\tilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$. In turn, it follows from [6, Proposition 8.26] that g is convex and from [6, Lemma 1.29] that it is lower semicontinuous and proper. \square

5 Integral functions

In this section we construct lower semicontinuous functions by using as an integrand a perspective function. First, let us extend and formalize the divergence model (1.3).

Proposition 5.1 *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let G be a separable real Hilbert space, and let $\varphi \in \Gamma_0(G)$. Set $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathbb{R})$ and $\mathcal{G} = L^2((\Omega, \mathcal{F}, \mu); G)$, and suppose that one of the following holds:*

- (i) $\mu(\Omega) < +\infty$.
- (ii) $\varphi \geq \varphi(0) = 0$.

For every $x \in \mathcal{H}$, set $\Omega_0(x) = \{\omega \in \Omega \mid x(\omega) = 0\}$ and $\Omega_+(x) = \{\omega \in \Omega \mid x(\omega) > 0\}$. Define

$$\Phi: \mathcal{H} \oplus \mathcal{G} \rightarrow]-\infty, +\infty] : (x, y) \mapsto \begin{cases} \int_{\Omega_0(x)} (\text{rec } \varphi)(y(\omega)) \mu(d\omega) + \int_{\Omega_+(x)} x(\omega) \varphi\left(\frac{y(\omega)}{x(\omega)}\right) \mu(d\omega), \\ \\ +\infty, \end{cases} \quad \text{if } \begin{cases} x \geq 0 \text{ a.e.} \\ (\text{rec } \varphi)(y)1_{\Omega_0(x)} + x\varphi(y/x)1_{\Omega_+(x)} \in L^1((\Omega, \mathcal{F}, \mu); \mathbb{R}); \\ \text{otherwise.} \end{cases} \quad (5.1)$$

Then $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$.

Proof. It follows from Proposition 2.2(ii) that $\tilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus G)$. Furthermore, we derive from (2.2) and (5.1) that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{G}) \quad \Phi(x, y) = \int_{\Omega} \tilde{\varphi}(x(\omega), y(\omega)) \mu(d\omega). \quad (5.2)$$

In turn, [6, Proposition 9.32] yields $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$. \square

Proposition 5.2 *Let Ω be a nonempty open subset of \mathbb{R}^N and let \mathcal{H} be the Sobolev space $H^1(\Omega)$, i.e., $\mathcal{H} = \{x \in L^2(\Omega) \mid \nabla x \in (L^2(\Omega))^N\}$. For every $x \in \mathcal{H}$, set $\Omega_-(x) = \{t \in \Omega \mid x(t) < 0\}$, $\Omega_0(x) = \{t \in \Omega \mid x(t) = 0\}$, and $\Omega_+(x) = \{t \in \Omega \mid x(t) > 0\}$. Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ be such that $\varphi \geq \varphi(0) = 0$, and define*

$$f: \mathcal{H} \rightarrow]-\infty, +\infty] \quad x \mapsto \begin{cases} \int_{\Omega_0(x)} (\text{rec } \varphi)(\nabla x(t)) dt + \int_{\Omega_+(x)} x(t) \varphi\left(\frac{\nabla x(t)}{x(t)}\right) dt, & \text{if } x \geq 0 \text{ a.e.;} \\ +\infty, & \text{otherwise.} \end{cases} \quad (5.3)$$

Then $f \in \Gamma_0(\mathcal{H})$.

Proof. Set $\mathcal{G} = (L^2(\Omega))^N$ and $\mathbf{G} = \mathbb{R}^N$, define Φ as in (5.1), where $(\Omega, \mathcal{F}, \mu)$ is the standard Lebesgue measure space, and let $L: \mathcal{H} \rightarrow \mathcal{H} \oplus \mathcal{G}: x \mapsto (x, \nabla x)$. Then $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$ by Proposition 5.1(ii). On the other hand, since $\nabla: \mathcal{H} \rightarrow \mathcal{G}$ is bounded, L is linear and continuous. Finally, since $f(0) = 0$, we conclude that $f = \Phi \circ L \in \Gamma_0(\mathcal{H})$. \square

The next examples recover two classical functions that have been extensively used in statistics (Fisher information) and in image recovery (total variation).

Example 5.3 Consider the setting of Proposition 5.2.

- (i) By choosing the supercoercive function $\varphi = \|\cdot\|_2^2$, we infer that the Fisher information

$$f: H^1(\Omega) \rightarrow]-\infty, +\infty]$$

$$x \mapsto \begin{cases} \int_{\Omega_+(x)} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt, & \text{if } \begin{cases} x \geq 0 \text{ a.e.} \\ [x = 0 \Rightarrow \nabla x = 0] \text{ a.e.} \end{cases} \\ +\infty, & \text{otherwise} \end{cases} \quad (5.4)$$

is in $\Gamma_0(H^1(\Omega))$. The convexity properties of (1.2) over the subspace of strictly positive 1-dimensional smooth densities were apparently first discussed in [21].

- (ii) By choosing the positively homogeneous function $\varphi = \|\cdot\|_2$, we infer that the total variation function

$$f: H^1(\Omega) \rightarrow]-\infty, +\infty]$$

$$x \mapsto \begin{cases} \int_{\Omega} \|\nabla x(t)\|_2 dt, & \text{if } x \geq 0 \text{ a.e.;} \\ +\infty, & \text{otherwise} \end{cases} \quad (5.5)$$

is in $\Gamma_0(H^1(\Omega))$.

We can also derive from Proposition 5.1 lower semicontinuous versions of a variety of standard divergences in the continuous and discrete cases. In the former, the underlying measure space is the Lebesgue measure space. The latter is illustrated below.

Example 5.4 Let N be a strictly positive integer, set $I = \{1, \dots, N\}$, and let $\phi \in \Gamma_0(\mathbb{R})$. Given $x = (\xi_i)_{i \in I} \in \mathbb{R}^N$ and $y = (\eta_i)_{i \in I} \in \mathbb{R}^N$, set $I_-(x) = \{i \in I \mid \xi_i < 0\}$, $I_0(x) = \{i \in I \mid \xi_i = 0\}$, $I_+(x) = \{i \in I \mid \xi_i > 0\}$, and

$$d_\phi(x, y) = \begin{cases} \sum_{i \in I_0(x)} (\text{rec } \phi)(\eta_i) + \sum_{i \in I_+(x)} \xi_i \phi(\eta_i / \xi_i), & \text{if } I_-(x) = \emptyset; \\ +\infty, & \text{if } I_-(x) \neq \emptyset. \end{cases} \quad (5.6)$$

Then $d_\phi \in \Gamma_0(\mathbb{R}^{2N})$. Indeed, this is a special case of Proposition 5.1(i), where $\Omega = \{1, \dots, N\}$, $\mathcal{F} = 2^\Omega$, μ is the counting measure (hence $\mathcal{H} = \mathcal{G} = \mathbb{R}^N$), $\varphi = \phi$, and $\mathbf{G} = \mathbb{R}$. For instance, consider

$$\phi: \mathbb{R} \rightarrow]-\infty, +\infty]: t \mapsto \begin{cases} t \ln t, & \text{if } t > 0; \\ 0, & \text{if } t = 0; \\ +\infty, & \text{if } t < 0. \end{cases} \quad (5.7)$$

Then $\text{rec } \phi = \iota_{\{0\}}$ and, if we set $I_0(x, y) = \{i \in I \mid \xi_i = 0 \text{ and } \eta_i \neq 0\}$, $d_\phi(x, y)$ is the Kullback-Leibler divergence between x and y , i.e.,

$$d_\phi(x, y) = \begin{cases} \sum_{i \in I_+(x)} \eta_i \ln(\eta_i / \xi_i), & \text{if } I_-(x) \cup I_0(x, y) = \emptyset; \\ +\infty, & \text{otherwise.} \end{cases} \quad (5.8)$$

This notion is central in statistics and in information theory. Another noteworthy family of discrete divergences is obtained by replacing (5.7) by

$$\phi: \mathbb{R} \rightarrow]-\infty, +\infty] : t \mapsto \begin{cases} |t^{1/p} - 1|^p, & \text{if } t \geq 0; \\ +\infty, & \text{if } t < 0, \end{cases} \quad \text{where } p \in [1, +\infty[. \quad (5.9)$$

In this case $\text{rec } \phi = \sigma_{]-\infty, 1]}$ and, if we set $I_0(x, y) = \{i \in I \mid \xi_i = 0 \text{ and } \eta_i < 0\}$, (5.6) becomes

$$d_\phi(x, y) = \begin{cases} \sum_{i \in I_0(x)} \eta_i + \sum_{i \in I_+(x)} |\eta_i^{1/p} - \xi_i^{1/p}|^p, & \text{if } I_-(x) \cup I_0(x, y) = \emptyset; \\ +\infty, & \text{otherwise.} \end{cases} \quad (5.10)$$

We recover the Kolmogorov variational divergence for $p = 1$ and the Hellinger divergence for $p = 2$.

Acknowledgement. The work of P. L. Combettes was partially supported by the CNRS MASTODONS project under grant 2016TABASCO.

References

- [1] M. Akian, S. Gaubert, and A. Hochart, Minimax representation of nonexpansive functions and application to zero-sum recursive games, 2016. <http://arxiv.org/abs/1605.04518>
- [2] S. M. Ali and S. D. Silvey, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc.*, vol. B28, pp. 131–142, 1966.
- [3] J. J. Alibert, G. Bouchitté, I. Fragalà, and I. Lucardesi, A nonstandard free boundary problem arising in the shape optimization of thin torsion rods, *Interfaces Free Bound.*, vol. 15, pp. 95–119, 2013.
- [4] M. Basseville, Distance measures for signal processing and pattern recognition, *Signal Processing*, vol. 18, pp. 349–369, 1989.
- [5] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Comm. Contemp. Math.*, vol. 3, pp. 615–647, 2001.
- [6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
- [7] A. Ben-Tal, A. Ben-Israel, and M. Teboulle, Certainty equivalents and information measures: Duality and extremal principles, *J. Math. Anal. Appl.*, vol. 157, pp. 211–236, 1991.
- [8] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.*, vol. 84, pp. 375–393, 2000.
- [9] J.-F. Bercher, Some properties of generalized Fisher information in the context of nonextensive thermostatics, *Physica A*, vol. 392, pp. 3140–3154, 2013.

- [10] A. Berline and I. Vajda, Selection rules based on divergences, *Statistics*, vol. 45, pp. 479–495, 2011.
- [11] J. Bien, I. Gaynanova, J. Lederer, and C. L. Müller, Non-convex global minimization and false discovery rate control for the TREX, 2016. <http://arxiv.org/abs/1604.06815>
- [12] D. E. Boekee, An extension of the Fisher information measure, in: I. Csiszár and P. Elias (eds.), *Topics in Information Theory*, János Bolyai Mathematical Society, vol. 16, pp. 113–123. North-Holland, Keszthely, Hungary, 1977.
- [13] J. M. Borwein, A. S. Lewis, M. N. Limber, and D. Noll, Maximum entropy reconstruction using derivative information, part 2: Computational results, *Numer. Math.*, vol. 69, pp. 243–256, 1995.
- [14] J. M. Borwein, A. S. Lewis, and D. Noll, Maximum entropy reconstruction using derivative information, part 1: Fisher information and convex duality, *Math. Oper. Res.*, vol. 21, pp. 442–468, 1996.
- [15] G. Bouchitté, I. Fragalà, I. Lucardesi, and P. Seppecher, Optimal thin torsion rods and Cheeger sets, *SIAM J. Math. Anal.*, vol. 44, pp. 483–512, 2012.
- [16] M. L. Bougeard, Connection between some statistical estimation criteria, lower-C2 functions and Moreau-Yosida approximates, in: *Bulletin International Statistical Institute*, 47th session, contributed papers, vol. 1, 159–160, 1989.
- [17] M. L. Bougeard and C. D. Caquineau, Parallel proximal decomposition algorithms for robust estimation, *Ann. Oper. Res.*, vol. 90, pp. 247–270, 1999.
- [18] L. Brasco, G. Buttazzo, and F. Santambrogio, A Benamou-Brenier approach to branched transport, *SIAM J. Math. Anal.*, vol. 43, pp. 1023–1040, 2011.
- [19] S. Ceria and J. Soares, Convex programming for disjunctive convex optimization, *Math. Program.*, vol. A86, pp. 595–614, 1999.
- [20] G. Choquet, *Topologie*. Masson, Paris, 1964 (English translation: *Topology*. Academic Press, New York, 1966).
- [21] M. L. Cohen, The Fisher information and convexity, *IEEE Trans. Inform. Theory*, vol. 14, pp. 591–592, 1968.
- [22] P. L. Combettes and C. L. Müller, Perspective functions: Proximal calculus and applications in high-dimensional statistics, 2016. <https://arxiv.org/abs/1610.01478>
- [23] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.
- [24] B. Dacorogna and P. Maréchal, The role of perspective functions in convexity, polyconvexity, rank-one convexity and separate convexity, *J. Convex Anal.*, vol. 15, pp. 271–284, 2008.
- [25] R. A. Fisher, Theory of statistical estimation, *Proc. Cambridge. Philos. Soc.*, vol. 22, pp. 700–725, 1925.
- [26] J. H. Fitschen, F. Laus, and G. Steidl, Transport between RGB images motivated by dynamic optimal transport, *J. Math. Imaging Vis.*, published online 2016-04-06.
- [27] B. R. Frieden and R. A. Gatenby (eds.), *Exploratory Data Analysis Using Fisher Information*. Springer, New York, 2007.
- [28] H. Hijazi, P. Bonami, G. Cornuéjols, and A. Ouorou, Mixed-integer nonlinear programs featuring “on/off” constraints, *Comput. Optim. Appl.*, vol. 52, pp. 537–558, 2012.
- [29] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*. Springer-Verlag, New York, 1993.
- [30] J.-B. Hiriart-Urruty and J.-E. Martínez-Legaz, Convex solutions of a functional equation arising in information theory, *J. Math. Anal. Appl.*, vol. 328, pp. 1309–1320, 2007.
- [31] P. J. Huber, Robust estimation of a location parameter, *Ann. Stat.*, vol. 35, pp. 73–101, 1964.

- [32] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. Wiley, New York, 2009.
- [33] M. N. Jung, C. Kirches, and S. Sager, On perspective functions and vanishing constraints in mixed-integer nonlinear optimal control, in: *Facets of Combinatorial Optimization*, pp. 387–417. Springer, Heidelberg, 2013.
- [34] J. Lederer and C. L. Müller, Don’t fall for tuning parameters: Tuning-free variable selection in high dimensions with the TREX, *Proc. Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 2729–2735. AAAI Press, Austin, 2015.
- [35] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. Inform. Theory*, vol. 52, pp. 4394–4412, 2006.
- [36] C. A. Micchelli, J. M. Morales, and M. Pontil, Regularizers for structured sparsity, *Adv. Comput. Math.*, vol. 38, pp. 455–489, 2013.
- [37] N. Moehle and S. Boyd, A perspective-based convex relaxation for switched-affine optimal control, *Systems Control Lett.*, vol. 86, pp. 34–40, 2015.
- [38] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.*, vol. 255, pp. 2897–2899, 1962.
- [39] E. Ndiaye, O. Fercoq, A. Gramfort, V. Leclère, and J. Salmon, Efficient smoothed concomitant lasso estimation for high dimensional regression, <http://arxiv.org/abs/1606.02702>, 2016.
- [40] M. Nikolova and M. K. Ng, Analysis of half-quadratic minimization methods for signal and image recovery, *SIAM J. Sci. Comput.*, vol. 27, pp. 937–966, 2005.
- [41] D. Noll, Reconstruction with noisy data: An approach via eigenvalue optimization, *SIAM J. Optim.*, vol. 8, pp. 82–104, 1998.
- [42] A. B. Owen, A robust hybrid of lasso and ridge regression, *Contemp. Math.*, vol. 443, pp. 59–71, 2007.
- [43] N. Papadakis, G. Peyré, and E. Oudet, Optimal transport with proximal splitting, *SIAM J. Imaging Sci.*, vol. 7, pp. 212–238, 2014.
- [44] L. Pardo, *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC, Boca Raton, FL, 2006.
- [45] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [46] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer, New York, 2000.
- [47] C. Villani, Fisher information estimates for Boltzmann’s collision operator, *J. Math. Pures. Appl.*, vol. 77, pp. 821–837, 1998.
- [48] C. Zach and M. Pollefeys, Practical methods for convex multi-view reconstruction, *Lecture Notes in Comput. Sci.*, vol. 6314, pp. 354–367, 2010.
- [49] C. Zălinescu, *Convex Analysis in General Vector Spaces*. World Scientific, River Edge, NJ, 2002.